A Survey of Different Enhancements in Weighted PageRank Algorithm

Chintan Agravat
Department of Computer Engineering
L.D. College of Engineering Gujarat Technological University Ahmedabad, Gujarat, India

Abstract—Web pages are increasing day by day. Due to that problem arises how to give better search result for a given search engine query. Here it is shown a comparison between various page rank algorithms used for displaying Quality results for various search engine queries.

Key words: Ranker, WUM, WSM, WCM, QI, WPR

I. INTRODUCTION

Here the basic search engine architecture is shown from the given architecture. We clearly see that how the Search engine uses different steps to execute query. Here as shown in figure ranker is there. It uses different ranking algorithm for Quality Results. An efficient ranking of query words has a major role in efficient searching for query words [2]. Ranker to calculate scores for the matched documents according to their relevancies to the input query (L1 ranking), and select the top k Web pages to recalculate the scores with a more complex function (L2 ranking). The ranking functions are usually the combination of many Information Retrieval (IR) document features, which includes some query dependent dynamic and some query independent static features [1]. Page rank is static query independent.

In Section II -related work, III-comparison of ranking algorithms, IV-conclusion.

II. RELATED WORK

Web Content Mining: Web content mining deals with extracting information from the content data that is present in the web page to be conveyed to the user. The content of web pages can be text, image, audio, video, etc.[3]

Web Structure Mining: Web Structure Mining is the process of discovering Structure information in the web. This is used to analyse the link Structure of the web it uses two structures hyperlink and Document [3].

Web Usage Mining: Web Usage Mining is the process of discovering information from the usage data of the web pages.
Different variations of page rank algorithm falls under different categories of web mining. According to methods used to construct algorithm. It may use combinations of web mining methods. Generally these algorithm falls under web structure mining. As link is main concern in these algorithms

A. Weighted Page Rank Algorithm

This algorithm was proposed by Wenpu Xing and Ali Ghorbani which is an extension of PageRank algorithm. This Algorithm assigns rank values to pages according to their importance rather than dividing it evenly. The importance is assigned in terms of weight values to incoming and outgoing links. This is denoted as Win (m,n) and Wout (m,n) respectively.

\[ W_{in}^{m,n} = \frac{l_n}{\sum_{p \in R(m)} l_p} \]  
\[ (1) \]

In is number of incoming links of page n, Ip is number of incoming links of page p, R(m) is the reference page list of page m. Wout (m,n) is the weight of link(m,n). It is calculated on the basis of the number of outgoing links of page n and the number of outgoing links of all the reference pages of page m.

\[ W_{out}^{m,n} = \frac{o_n}{\sum_{p \in R(m)} o_p} \]  
\[ (2) \]

On is number of outgoing links of page n, Op is number of outgoing links of page p, and then the weighted PageRank is given by formula:

\[ WPR(n) = (1 - d) + d \sum_{m \in B(n)} WPR(m) W_{out}^{m,n} W_{in}^{n,m} \]  
\[ (3) \]

The complexity of Page Rank Algorithm is < \log n[5].

1) Limitations
- Certain pages which are not relevant to the query are also included in the result set because of their popularity. For example home pages are also ranked higher because of their greater number of inlinks and outlinks
- Topic Drift can occur.
- Weightage distributed proportionately to pages based on popularity.
- Query Independent [3].

B. Weighted PageRank Algorithm Based on no. of visits of the links of web page. [4]

This Algorithm is proposed by Neelam Tyagi and Simple Sharna in 2012. It is obvious depending upon inward and outward links Page Rank is assigned in Weight Based Page Ranking Algorithm. Based on the Number of Outlinks and Number of visitors of that link this algorithm assigns higher page rank value popularity of the webpage is not considered for this algorithm users behaviour and trend of search can be evaluated based on Visits of particular link.

\[ WPR_{vol}(u) = (1 - d) + d \sum_{v \in B(u)} \frac{WPR_{vol}(v) W_{in}^v}{TL(v)} \]  
\[ (4) \]

\( d \)= dampening Factor
\( u=\text{webpage} \)
\( B(u)=\text{set of pages referring to u.} \)
\( WPR_{vol}(u) \) and \( WPR_{vol}(v) \) are Ranks assigned to u and v.
\( L_v \) is Number of Visits of Link from v to u.
\( TL(v) \) denotes total Number of visits of all Links which are on the v.

C. On the Improvement of Weighted Page Content Rank [5]

This Algorithm was proposed by Seifedine Kadry and Ali Kalakech in 2013. This Algorithm works on the concept first it calculates WPR. Then it calculates Content Weight Factor. For these Pages. It is mainly depended on the query terms. It calculates relevance of the page based upon two factors.

- F(pi)=gives the freqery of occurring query text.
- CWF(pi)=GPA(f(pi)) is core for proposed algorithm.
- Matrix sized x*y where,
- x=number of pages.
- y=Word count in given query each column gives frequency f(y).

<table>
<thead>
<tr>
<th>F(1)</th>
<th>F(1)</th>
<th>F(1)</th>
<th>F(1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>......</td>
<td>......</td>
<td>......</td>
<td>......</td>
</tr>
<tr>
<td>F(y-1)</td>
<td>F(y-1)</td>
<td>F(y-1)</td>
<td>F(y-1)</td>
</tr>
<tr>
<td>F(y)</td>
<td>F(y)</td>
<td>F(y)</td>
<td>F(y)</td>
</tr>
<tr>
<td>P1</td>
<td>P2</td>
<td>......</td>
<td>P(X)</td>
</tr>
</tbody>
</table>

D. Improvement in Weighted Page Rank based on Visits of Links (VOL) Algorithm [6]

It was discovered by Sachin Gupta and Pallvi Mahajan in 2014. This Algorithm uses the user attention time means how much time user spent on particular page. It is an improvement to the WPR-visitor of links Algorithm. Other Features are same

\[ WPR_{vol}(m) = d + (1 - d) * UAT(m) * \sum_{n \in B(m)} \frac{WPR_{vol}(n) W_{in}^n}{TL(n)} \]  
\[ (5) \]
A Survey of Different Enhancements in Weighted PageRank Algorithm

- \( d \) = dampening Factor
- \( m \) = webpage
- \( B(m) \) = set of pages referring to u.
- \( WPR_{out}(m) \) and \( WPR_{out}(n) \) are Ranks assigned to m and n.
- \( Lm \) is Number of Visits of Link from n to m.
- \( UAT(m) \) is time spent by user on that page.

E. Weighted Page Rank Algorithm Based on Number of Visits of Links of Web Pages in Time Duration [7]

The Algorithm was given by Anushree Gambhir, Arushi Goyal, Sumit singh and Yatharth Srivastav in 2014. It is improved from WPR-Based on Number of visitors of link. Improvement in this algorithm is that it considers Number of visitors for a particular time period that’s why new releasing website have equal opportunity to get higher place in search result.

\[
WPR_{vol}(p) = \left[ (1 - d) + d \sum_{q \in R(p)} \frac{L_p \cdot WPR_{vol}(q) \cdot W_{in}(p, q)}{TL(q)} \right] \ast t + k
\]  

- \( t \) represents time (year in this scenario) under consideration.
- \( k \) constant
- \( p \) = web page,
- \( d \) = dampening factor.
- \( WPR(p) \) and \( WPR(q) \) are rank values of page \( p \) and \( q \) respectively.
- \( L(p) \) = number of visits of link which is pointing page \( p \) from \( q \).
- \( TL(q) \) = total number of visit of all links present on \( q \).

F. Agent Based Weighted Page Ranking Algorithm for Web Content Information Retrieval[8]

It was being proposed by Nagappan V.K. and Dr. P. Elango in 2015. It mainly uses probabilities of weight and content’s weight. Summation of these two values is taken as multiplying factor to the weighted page ranking formula.

Content weight (CW) = \( u/v \)

Probability Weight (PW) = \( m/n \)

- \( u \) = Sum of occurring of strings in SMP (Set of maximum possible occurring).
- \( v \) = Summation of frequencies of all meaningful word strings of Query (Q).
- \( m \) = Number of query terms in the webpage.
- \( n \) = Number of entire query terms of Q whereas stop words are ignored.

\[
PR(i) = (1 - d) + d \sum_{j \in E(i)} PR(j) \cdot W_{in}(i, j) \cdot W_{out}(i, j) \ast (CW + PW)
\]  

- \( PR(i) \) = rank of page \( i \)
- \( R(j) \) = Pages referencing towards \( j \).
- \( d \) = dampening value.
- \( W_{in}(i, j) \) = inwards weight of link \( (i, j) \).
- \( W_{out}(i, j) \) = outwards weight of link \( (i, j) \).

G. Weighted Page Rank Algorithm based on In-Out Weight of Webpages[9]

This algorithm is given by B jaganathan and kalyani desikan in 2015. It mainly follows weighted pageRank Property but it also calculates the matrix for given web graph. Using that adjacency matrix it calculates the rank for the page in consideration according to given formulas.

\[
m_{(i,j)} = \begin{cases} 
1 & \text{if } i \text{ and } j \text{ is pointing/links to } v \text{ j} \\
0 & \text{in other cases}
\end{cases}
\]

According to this formula created matrix is called adjacency matrix AM(G). Weights of in and out link are calculated using weighted pagerank formula. Matrices for in and out link weights are created. And Summation of both matrices is taken as a factor of multiplication to original algorithm.

Weight matrix formula,

\[
W_{(v, u)}(G) = W_{in}^{(v, u)} + W_{out}^{(v, u)}
\]  

PageRank calculation form iterations,

\[
PR_m = (1 - d) + d \ast W(G) \ast AM(G) \ast PR_{m-1}
\]  

Here in this algorithm iterative approach is used to calculate final rank.

H. A New Ranking Algorithm for Search Engine: Content’s weight based page ranking[10]

It is given by charanjit singh, vijay laxmi and arvinder singh kang in 2016. It will not assign rank in equal proportion among all pages. First it will calculate the frequency of query term in each URL. According to that it will assign rank initially to all pages. Then rank will be assigned according to WPR. So, the Pages having more Number of query terms will get higher priority. So, two algorithms applied to calculate final result.

It was proposed by Balwant kumar and Dharmender kumar in 2017. WPR-VOL is extended in this algorithm. WPR-VOL considers only inbound links to calculate visitor in this algorithm inbound as well as outbound link is also taken into consideration. Formula is same as WPR-VOL but while calculating weight it considers outlinks as well.

\[
\text{NWPR}_\text{vol}(u) = (1 - d) + d \sum_{v \in R(u)} \frac{L_v \text{NWPR}_\text{vol}(v) W_v(u, v)}{\text{TL}(v)}
\]

It is formula to calculate NWPR for given pages.

III. *Conclusion*

Different Enhancements of weighted PageRank Algorithm gives better result than older Algorithm in some cases. They are giving these results at the cost of higher time complexity. Other factors are added to improve quality of result.

REFERENCES


